Search for $B_{s,d} \to \mu^+\mu^-$ with CDF II

Walter Hopkins

Cornell University

April 26$^{\text{th}}$ 2012

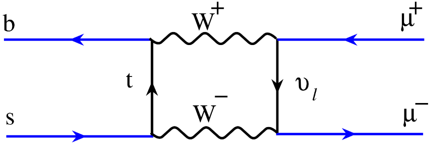
D. Glenzinski, M. Herndon, T. Kamon, D. Kong, V. Krutelyov, C. J. Lin, J. Thom, S. Uozumi

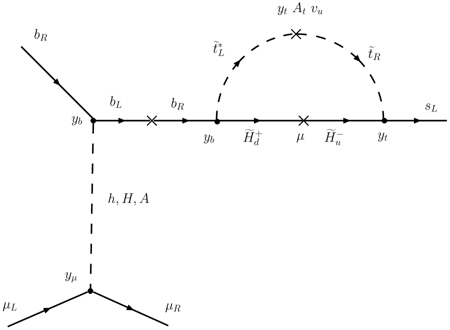
Argonne National Laboratory Seminar

PRL 107, 191801 (2011)
PRD in preparation

Introduction

- $B_s \to \mu^+\mu^-$ can only occur through higher order FCNC diagrams in Standard Model (SM)
- Suppressed by the GIM Mechanism and helicity
- SM predicts very low rate with little SM background:
  $\mathcal{B}(B_s \to \mu^+\mu^-) = (3.2 \pm 0.2) \times 10^{-9}$
  $\mathcal{B}(B_d \to \mu^+\mu^-) = (1.0 \pm 0.1) \times 10^{-10}$
  E.Gamiz et al. (HPQCD Collaboration), A.J. Buras et al.
- New Physics models predict enhancement
- Clean experimental signature





Motivation: BSM prediction

- Large new contributions from models with new operators
- Modest enhancements without new operators
- Ratio of $\mathcal{B}(B_s \to \mu^+\mu^-)$ and $\mathcal{B}(B_d \to \mu^+\mu^-)$ is important to discriminate amongst BSM models
- Correlation between CP violating phase in $B_s \to J/\Psi\phi$ and $\mathcal{B}(B_s \to \mu^+\mu^-)$

Model	$\mathcal{B}(B_{s,d} \to \mu^+\mu^-)$ Enhance
MFV	1000%
CMFV	20%
LHT	30%
RS	10%
4G	250%
AC	1000%
RVV	1000%

Table: Maximal enhancements for $\mathcal{B}(B_{s,d} \to \mu^+\mu^-)$ from different theoretical NP models. SUSY Models: MFV=Minimal Flavor Violation; AC=Agashe, Carone; RVV=Ross, Velaso-Sevilla, Vives. Plenary talk, A.Buras, Beauty 2011

Powerful tool in NP model discrimination

Analysis Description

Simple Analysis

- 2 Muons low p_T muons ($p_T < 15$ GeV/c)
- Identify methods of suppressing background and keep signal
- Look for bump in di-muon mass distribution

Analysis Strategy

- Blind ourselves to di-muon signal mass region
- Use mass sidebands to estimate dominant background in signal region
- Optimize selection criteria a priori
- Build confidence in background estimates by employing same methods on control regions
- Unblind and perform statistical analysis of result

Analysis Properties and Techniques

Analysis Properties

- b Physics analysis, good MC modeling of b and c hadrons
- CDF is well understood detector
- Large data set ($\sim 10\text{fb}^{-1}$)
- Previous iterations of the analysis
- Mature calibrations

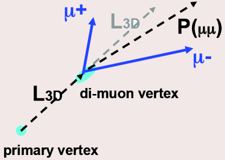
Analysis Techniques

- Use normalization to measure $\mathcal{B}(B_s \to \mu^+\mu^-)$
- Multi-variate analysis
- Control regions for background check
- Statistical interpretation: CLs limits, p-values, $\Delta\chi^2$ fit

Signal vs. Background

Signal Properties

- Final state fully reconstructed
- B_s is long lived ($c\tau =\sim 450\mu$m)
- B fragmentation is hard: few additional tracks

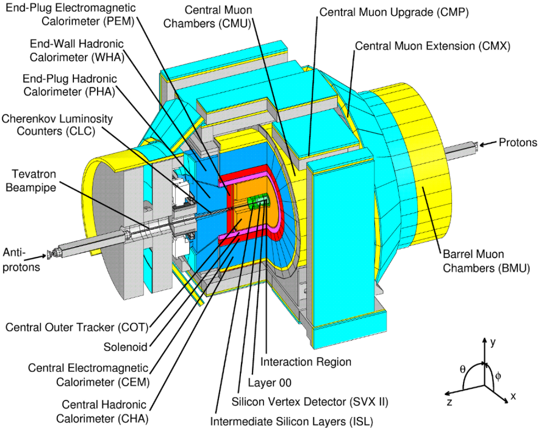


Background contributions & characteristics

- Sequential semi-leptonic decay: $b \rightarrow c\mu^- X \rightarrow \mu^+\mu^- X$
- Double semi-leptonic decay: $bb \rightarrow \mu^-\mu^+ X$
- Continuum $\mu^-\mu^+$
- μ + fake and fake+fake
    - Partially reconstructed
    - Softer
    - Short lived
    - Has more tracks
- $B \rightarrow h^+ h'^-$: peaking in signal region (h and h' are pions or kaons)

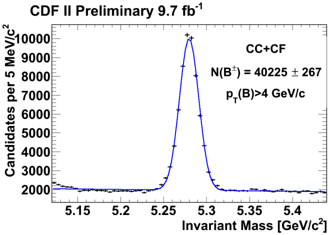
CDF II Detector

- Multi-purpose detector
- Silicon vertex detector close to beam line⇒ 35 μm vertex resolution
- Central Outer Tracker (COT) ⇒ multi-wire drift chamber
- Good Muon drift chambers: Central and some Forward (yellow and cyan)
- Use entire Run II data set



What do we measure?

- Measure rate of $B_s \to \mu^+\mu^-$ relative to $B^+ \to J/\Psi K^+$, $J/\Psi \to \mu^+\mu^-$
- Apply same selection to find $B^+ \to J/\Psi K^+$
- Systematic uncertainties will cancel in ratio



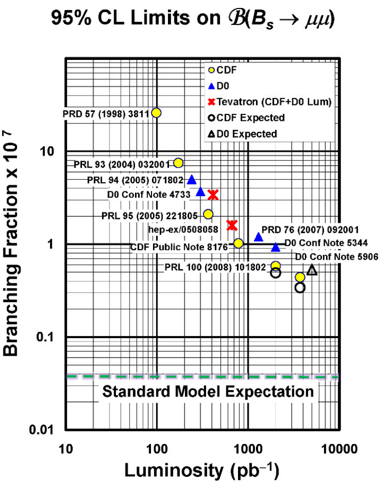
$$\mathcal{B}(B_s \to \mu^+\mu^-) = N_{B_s} \cdot \boxed{\frac{1}{N_{B^+}} \frac{\epsilon^{trig}_{B^+}}{\epsilon^{trig}_{B_s}}} \cdot \boxed{\frac{\epsilon^{reco}_{B^+}}{\epsilon^{reco}_{B_s}} \frac{\alpha_{B^+}}{\alpha_{B_s}} \frac{1}{\epsilon^{NN}_{B_s}}} \cdot \boxed{\frac{f_u}{f_s} \cdot \mathcal{B}(B^+ \to J/\Psi K^+ \to \mu^+\mu^- K^+)}$$

From Data, From MC, From PDG

$$\boxed{N_{B^+} \sim 2 \times 10^4,\ \frac{\epsilon^{trig}_{B^+}}{\epsilon^{trig}_{B_s}} \sim 1} \quad \boxed{\frac{\epsilon^{reco}_{B^+}}{\epsilon^{reco}_{B_s}} \sim 1,\ \frac{\alpha_{B^+}}{\alpha_{B_s}} \sim 0.5,\ \frac{1}{\epsilon^{NN}_{B_s}} \sim 1}$$

$$\boxed{\frac{f_u}{f_s} \sim 3,\ \mathcal{B}(B^+ \to J/\Psi K^+ \to \mu^+\mu^- K^+) \sim 5 \times 10^{-5}}$$

- Iterations of analysis before 2011
- CDF and D0 set upper limits on $B_s \to \mu^+\mu^-$
- Tightest limit from CDF with 3.7 fb^{-1} of data: $\mathcal{B}(B_s \to \mu^+\mu^-) < 4.3 \times 10^{-8}$ and $\mathcal{B}(B_d \to \mu^+\mu^-) < 7.6 \times 10^{-9}$ at 95% C.L.



Already greatly constrained NP parameter space
Closing on SM prediction (factor ~ 10)

Analysis Improvements after 3.7 fb^{-1} Iteration*

- Increased acceptance by including more forward detector regions ($\sim 7\%$)
- More than double the data set
- New multi-variate discriminant (NeuroBayes)
- New calibration for muon ID
- New background estimates
- Additional Statistical Interpretation

Event Selection

Trigger

Central-Central (CC)

- 2 Muons with $|\eta| < 0.6$
- $p_T > 1.5$ GeV$/c$ or $p_T > 3.0$ GeV$/c$ $\rightarrow$ Range out
- Muons must be separated by $\Delta\phi_{SL6} > 1.25°$ $\rightarrow$ tracking/muon stub granularity

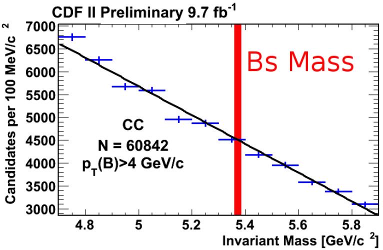
Central-Forward (CF)

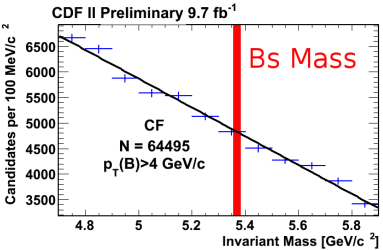
- Muon with $|\eta| < 0.6$ and muon with $0.6 < |\eta| < 1.0$
- $p_T > 2.0$ GeV$/c$ $\rightarrow$ Range out

- Opposite sign muons
- $|\Delta z_0| < 5$ cm $\rightarrow$ Should come from same source

Baseline Requirements

- $p_T(\mu) > 2.0(2.2)$ GeV$/c$ $\rightarrow$ **Rapidly changing trigger eff**
- $p_T(B_s) > 4.0$ GeV$/c$
- Hits in 3 layers of SVX $\rightarrow$ **Improved impact parameter resolution**
- Muon likelihood and dE/dx$\rightarrow$ **Kaon rejection**
- Vertex: Proper decay length, χ^2 of vertex, etc $\rightarrow$ **Reject short lived background**
- Invariant mass
- Isolation and Pointing angle $\rightarrow$ **Reject jets and short lived**
    - Isolation $= \frac{p_T(\mu\mu)}{\Sigma p_T(\text{other tracks}) + p_T(\mu\mu)}$ in R=1.0 $\eta - \phi$ cone.
    - Pointing angle = angle between di-muon momentum vector and vector pointing from primary to secondary vertex



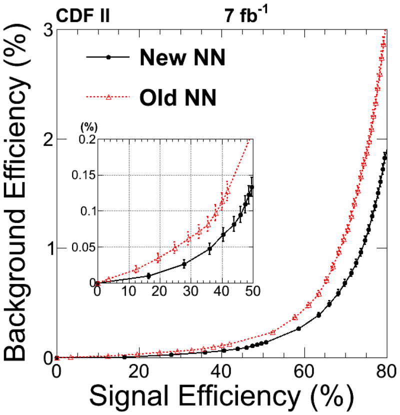


New Neural Network

- New 14-variable NN to increase S/B
  - Studied different input variables
- Carefully chose input variables to avoid bias in $M_{\mu\mu}$
- Twice the background rejection as old NN

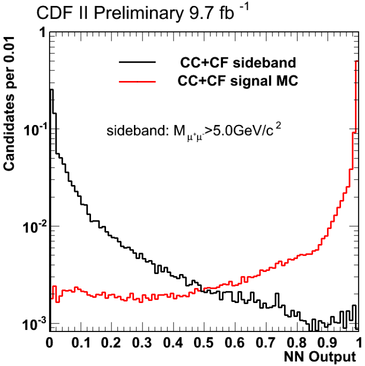
NN Input Variables

- λ (proper decay length)
- Isolation
- Pointing angle
- λ/σ_λ
- lower $p_T(\mu)$
- Secondary vertex χ^2
- Decay length (L_{3D})
- Transverse Decay length significance ($L_{xy}/\sigma_{L_{xy}}$)
- 2D Pointing angle
- Smaller impact parameter
- Larger impact parameter
- Smaller impact parameter significance
- Larger impact parameter significance
- $B_{s(d)}$ impact parameter



Neural Network Training

- Signal training sample: $B_s \to \mu^+\mu^-$ MC
- Background sample: di-muon mass sideband for combinatorial background rejection
- Separate NN's for CC and CF
- Investigated 20 input variables
- Excluded variables that caused di-muon mass/NN output correlation
    - Opening angle between muons
    - $p_T(B_s)$
- Final NN used 14 strongest separating variables
- Combined separation power into 1 output that ranges between 0 and 1



NN Signal Region

- Chose NN$>$ 0.7
- Divided region into 8 NN bins based on expected limit optimization

Signal Efficiencies

$$\mathcal{B}(B_s \to \mu^+\mu^-) = N_{B_s} \cdot \frac{1}{N_{B^+}} \boxed{\frac{\epsilon^{trig}_{B^+}}{\epsilon^{trig}_{B_s}} \cdot \frac{\epsilon^{reco}_{B^+}}{\epsilon^{reco}_{B_s}} \frac{\alpha_{B^+}}{\alpha_{B_s}} \frac{1}{\epsilon^{NN}_{B_s}}} \cdot \frac{f_u}{f_s} \cdot \mathcal{B}(B^+ \to J/\Psi K^+ \to \mu^+\mu^- K^+)$$

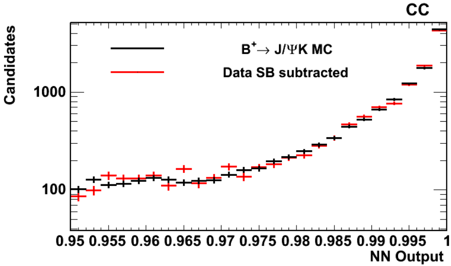
- Estimate total acceptance and efficiency
- Estimate separately for $B_s \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$
    - Kinematic differences in 2 and 3 body decays

Acceptance and Efficiency Broken Down

- α_{B_s}: Geometric and kinematic acceptance of trigger $\rightarrow$ estimated with MC
- $\epsilon^{trig}_{B_s}$: Trigger efficiency within acceptance $\rightarrow$ measured in data
- $\epsilon^{reco}_{B_s}$: Efficiency of baseline requirements for event passing trigger $\rightarrow$ estimated with data and MC
- ϵ^{NN}: Efficiency for each NN bin (B_s only) $\rightarrow$ estimated with MC

NN Efficiency*

- Estimated using $B_s \to \mu^+\mu^-$ MC
- Estimated for 8 NN bins
- Highest NN bin accounts for majority of sensitivity (46% efficiency)



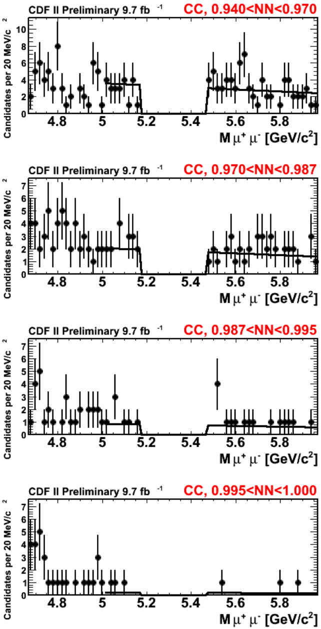
Systematics

- Apply NN to $B^+ \to J/\psi K^+$ MC and data and compare efficiencies
- Overall $\sim 5\%$ shift between MC and data
- Applied as systematic to highest NN bin
- Additional 4% systematic applied based on the iso and $p_T(B)$ MC mismodeling

Background Estimation

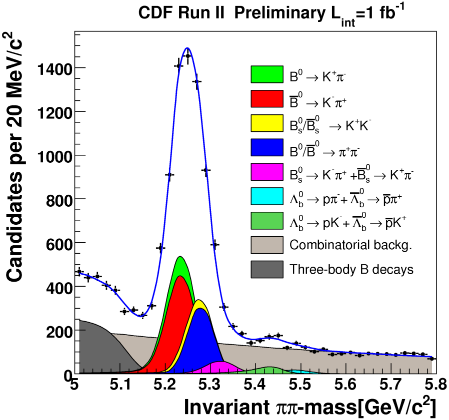
Combinatorial Background Estimates

- Exclude $M_{\mu^+\mu^-} < 5.0$ GeV$/c^2$ region, enhanced with $B \to \mu^+\mu^- X$ decays
- Fit first order polynomial to sidebands in each NN bin
- Estimate systematics due to shape uncertainty by fitting alternative function
  - Only for highest 3 NN bins
- Expect ~ 1 background event in CC channel and ~ 3 in CF channel for the highest NN bin



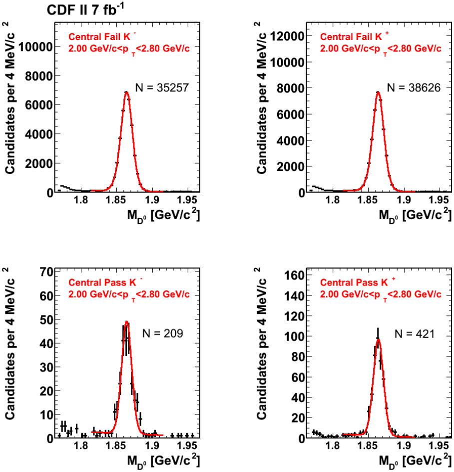
Peaking Background Estimates*

- Only significant peaking background is $B \to h^+ h'^-$ (h is hadron)
- Background from Λ_b decays much lower
  - Smaller production rates
  - Protons are significantly rejected by muon ID
- Estimated using MC and D^*-tagged $D^0 \to \pi^+ K^-$ data
  - MC for p_T and mass distributions
  - D^*-tagged $D^0 \to \pi^+ K^-$ data to asses rate at which pions/kaons fake muons
- More in B_d due to muon mass hypothesis



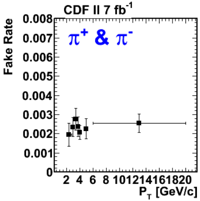
Data Set for Peaking Background*

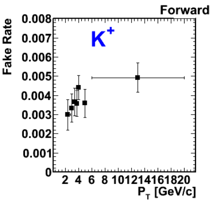
- Need to asses how often a kaon or pion passes muon reconstruction
- Use D^*-tagged $D^0 \to \pi^+ K^-$ data $\Rightarrow$ very pure samples of kaons and pions
- Sample collected with Two Track Trigger
- Numerator: Pass dE/dx and muon likelihood requirement
- Extract yield using Gaussian+pol fit

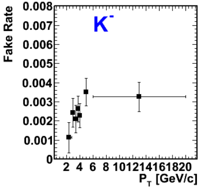


Fake Rate Parametrization*

- Separate fake rates for $\pi^{\pm}$, K^{+}, and K^{-}
- Parametrized in p_T ⇒ Higher momentum, more punch through
- Found inst. lumi dependence ⇒ estimated fake rate in 4 lumi bins
  - Fake rates changed by 20% to a factor of 3 due to lumi
- Applied fake rates as weights to D^* data and compared to actual number of fakes ⇒ Difference assigned as systematic.







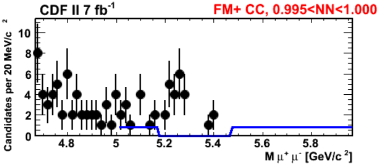
Background Estimate Check

- Check background estimates with background dominated control samples
  - Signal has two opposite sign muons with positive lifetime
  - Control samples have opposite sign negative lifetime, same-sign positive/negative lifetime, and reverse muon ID
  - Total of 64 samples
- Apply same background methods on control sample that we can unblind

NN cut	pred	obsv	prob(%)
0.700<NN<0.760	268.8±(14.3)	249	82.3
0.760<NN<0.850	320.8±(16.1)	282	95.1
0.850<NN<0.900	150.3±(9.9)	156	36.5
0.900<NN<0.940	146.2±(9.7)	158	23.0
0.940<NN<0.970	146.2±(9.7)	137	72.9
0.970<NN<0.987	100.4±(7.8)	98	58.3
0.987<NN<0.995	78.8±(6.8)	59	97.0
0.995<NN<1.000	41.2±(4.8)	42	47.2

$B \to hh$ Background Check*

- Used control sample with reversed muon ID cuts: enhanced in hadrons
  - Total of 16 samples
- Estimated fake rates for this sample using D^*-tagged for fake rate: Ratio of pions/kaons failing muon ID



Conclusion

- Checked combinatorial and peaking background estimates with control samples
- Good agreement between predicted and observed

Expected Sensitivity

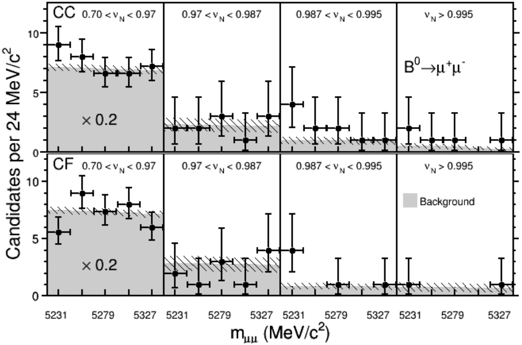
$B_s \to \mu^+\mu^-$ CC

NN Bin	ϵ_{NN}	B→hh Bkg	Total Bkg	Exp SM Signal
$0.700 < NN < 0.970$	20%	0.05	169.29±6.29	0.32±0.06
$0.970 < NN < 0.987$	8%	0.02	7.91±1.85	0.13±0.02
$0.987 < NN < 0.995$	12%	0.03	3.95±1.28	0.20±0.04
$0.995 < NN < 1.000$	46%	0.11	0.79±0.70	0.75±0.13

- $\sim$ 80% signal efficiency for NN
- Small contribution of peaking background compared to combinatorial (B_s)
- Expect $\sim$2 SM signal event (CC and CF)

B_d Results

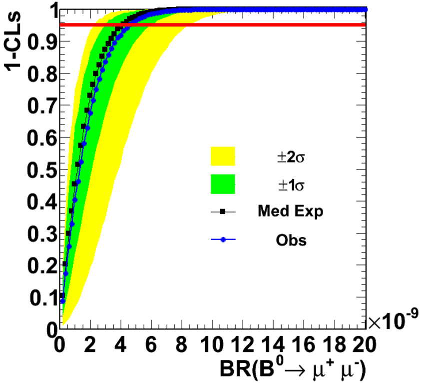
Comparison of observation and background estimates



- Five mass bins
- Five lowest NN bins combined
- Light gray: Background estimates, Hashed: Systematic errors on background
- Error bars on points: Poisson error on mean
- **No excess in B_d mass region**

CLs Bounds for B_d

- Generate background only pseudo data and $s+b$ pseudo data for many BR
- CLb = p-value using background-only pseudo data
- CLs+b = p-value using $s+b$ pseudo data
- CLs = CLs+b/CLb, exclude if 1-CLs>95%

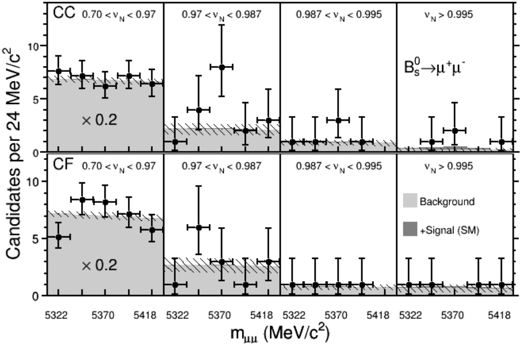


Results

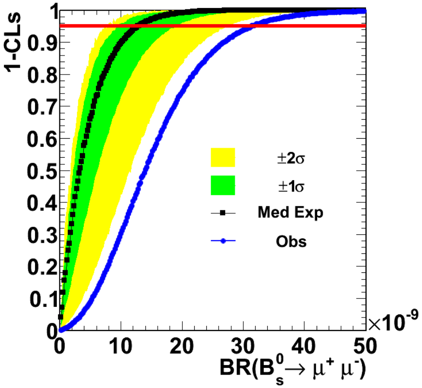
- $\boxed{\text{Observed: } \mathcal{B}(B_d \to \mu^+\mu^-) < 4.6 \times 10^{-9} \text{ @ 95\% C.L.}}$
- Expected $\mathcal{B}(B_d \to \mu^+\mu^-) < 4.0 \times 10^{-9}$ @ 95% C.L.
- SM prediction: $\mathcal{B}(B_d \to \mu^+\mu^-) = (1.0 \pm 0.1) \times 10^{-10}$

B_s Results

Comparison of observation and background estimates



- Dark gray: Expected SM signal
- **Excess over background-only in central region (the most sensitive)**

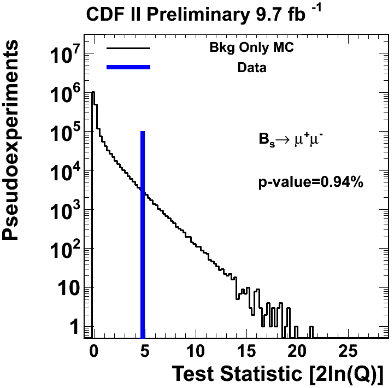


Results

- Observed: $\mathcal{B}(B_s \to \mu^+\mu^-)<3.1 \times 10^{-8}$ @ 95% C.L.
- Expected: $\mathcal{B}(B_s \to \mu^+\mu^-)<1.3 \times 10^{-8}$ @ 95% C.L. $\to > \mathbf{2}\sigma$ deviation
- SM Predicted: $\mathcal{B}(B_s \to \mu^+\mu^-) = (3.2 \pm 0.2) \times 10^{-9}$

p-Value Determination

- Construct likelihood function ($\mathcal{L}$): Product of 80 Poisson PDF's
- Considered 3 hypotheses
  - background-only, $\mathcal{L}(b)$, b from total background estimates
  - signal+background, $\mathcal{L}(s+b)$, s is floating
  - SM+background, $\mathcal{L}(\mathrm{SM}+b)$, from SM $\mathcal{B}(B_s \to \mu^+\mu^-)$
- Constructed log likelihood ratio: $2\ln(Q)$ with $Q = \frac{\mathcal{L}(s+b|data)}{\mathcal{L}(b|data)}$
- Generate pseudo-data while varying nuisance parameters
  - Systematics included as nuisance parameters modeled as Gaussians

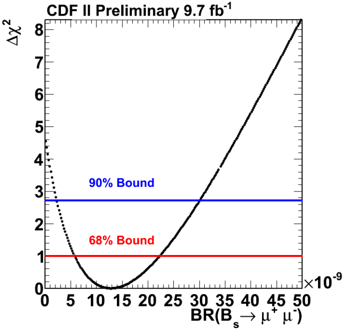


Results

- B_s bkg-only p-value: 0.94%
- B_s SM+bkg p-value: 7.1%
- (B_d bkg-only p-value: 41%)

B_s: Central Values, Bounds and P-Values

- Includes all systematics
- 90% Bound:
  $2.2 \times 10^{-9} < \mathcal{B}(B_s \to \mu^+\mu^-) < 3.0 \times 10^{-8}$
- Stable: No large deviation when only using subset of bins



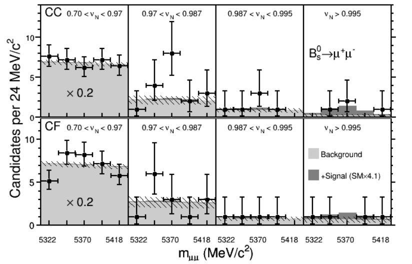
Summary of p-values and limits

	All Bins	2 Highest NN Bins
Best Fit ($\times 10^{-8}$)	$1.3^{+0.9}_{-0.7}$	$1.0^{+0.8}_{-0.6}$
90% Bounds ($\times 10^{-8}$)	$0.22 < \mathcal{B} < 3.0$	$0.08 < \mathcal{B} < 2.5$
Bkg Only p-value	0.94%	2.1%
SM+Bkg p-value	7.1%	22.5%

Third NN Bin Excess

Background Estimate Problem?

- Combinatorial Background Problem
  - B_d Uses same sideband as B_s $\Rightarrow$ No excess in B_d
- Peaking Background Problem
  - Only peaking background is $B \to hh$
  - 10x larger in B_d region
  - No excess in B_d $\Rightarrow$ good fake rates

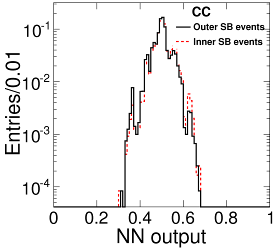


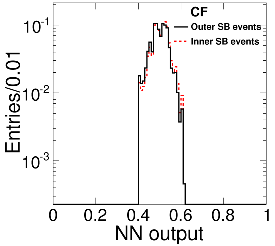
Neural Network Problem?

- $\boxed{\text{Mass bias? } \Rightarrow \text{ Many studies show no bias}}$
- Over-trained? $\Rightarrow$ Compared NN with different training sample, no difference
- Mismodels data? $\Rightarrow$ No difference between MC and data in normalization mode

NN Mass Correlation Studies: NN inner/outer SB training

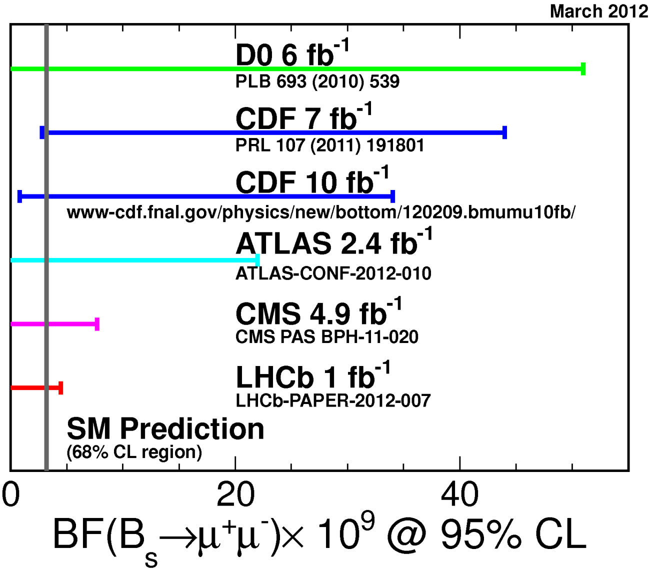
- Defined inner sideband close to signal region, and outer sideband
- Trained NN using inner sideband as signal and outer as background sample
- Inner and outer sideband regions are kinematically similar, di-muon mass is main difference
- Tests whether NN is selecting events based on di-muon mass





Conclusion: No difference in NN output for inner and outer after training
Mass bias unlikely to be cause of excess in 3rd NN bin of CC

Current Experimental Status



Summary

- New $B_s \to \mu^+\mu^-$ search with full CDF Run II data set
  - First CDF result using full Run II data set
- $0.8 \times 10^{-9} < \mathcal{B}(B_s \to \mu^+\mu^-) < 3.4 \times 10^{-8}$ @ 95% C.L.
  - First two sided bound
- $B_d \to \mu^+\mu^- < 4.6 \times 10^{-9}$ @ 95% C.L.
- PRD manuscript in preparation
- Many exciting new results from LHC
- Final CDF $B_s \to \mu^+\mu^-$ result
  - Tevatron Run II+CDF II+ingenuity provided 2 orders of magnitude improvement in sensitivity

x-Ray Beam Size Monitor

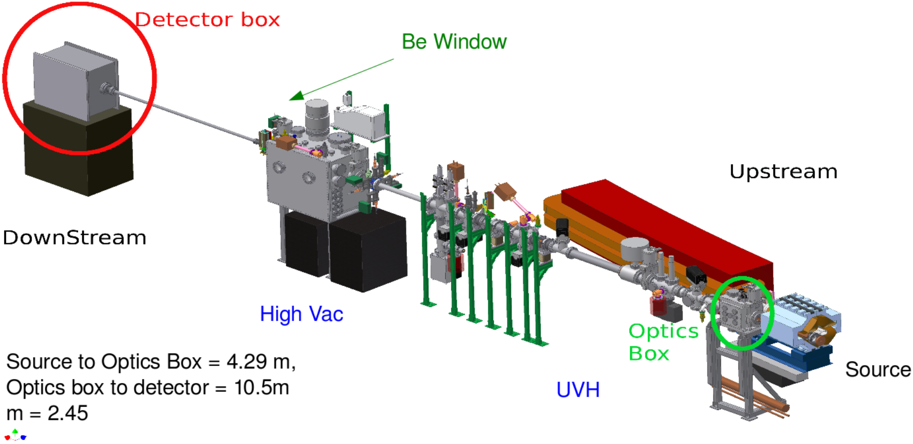
Intro to CESR-TA and xBSM

- Cornell Electron-positron Storage Ring Test Accelerator
- Test accelerator for cooling rings for ILC
- 14 ns bunch spacing

xBSM

- Need feedback from methods of reducing bunch size
- Measure bunch size every 14 ns using synchrotron x-rays

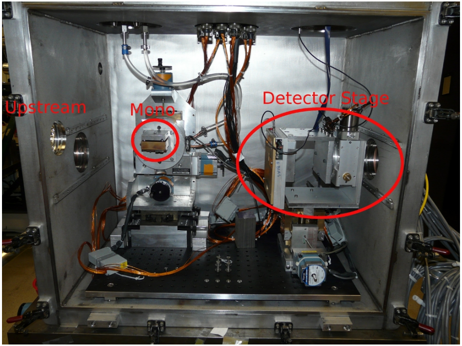
x-Ray Beam Line



Source to Optics Box = 4.29 m,
Optics box to detector = 10.5m
m = 2.45

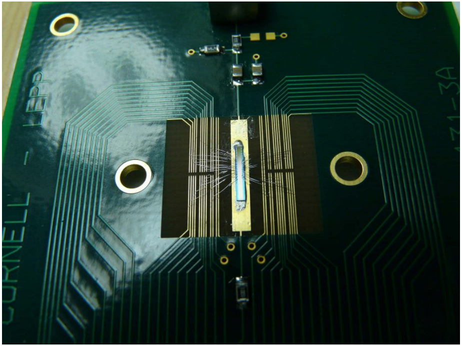
Beam Size Measurement

- 1D vertical beam size measurement
- Optics Elements: Pinhole, Fresnel Zone Plate, Coded Aperture
- Magnification of ~ 2.5
- Use monochromater to select small range in wave lengths
- Detector in vacuum, 10 m away from optics
    - Detector consists of 32 GaAs diodes in vertical orientation (measure 1D bunch size)
    - Can do integrated slow readout with each diode by moving motors
    - Snapshot readout: readout all 32 diodes every 14 ns



New Detector

- 32 diode GaAs
- New detector board (detector wire bonded)
    - Tested several detectors from vendors
    - Redesigned detector board for optimum wire bonding
- Commissioned new detector
    - Wrote new detector read out software
    - Read out mapping (diode number to physical location)
    - Gain calibration



xBSM Conclusion

Started with

- Integrated read out of one diode (moving the diode through beam)
- Positron read out only

After one year

- Selected 32 diode array and designed new detector board
- Automated diode mapping
- Gain calibration method established
- Successful beam profile for each 14 ns bunch
- Automatic 14 ns beam size reporting
- Started construction of electron beam size setup